

WHAT IS CLAIMED IS:

1. An identifier generating method comprising:

a canonicalization process step of subjecting document data to a canonicalization process to correct fluctuation of expression; and

an identifier generating step of generating an identifier uniquely specifying the document data or part thereof, based on all or part of the document data having been subjected to the canonicalization process in the canonicalization process step.

2. An identity determining method comprising:

a canonicalization process step of subjecting document data to a canonicalization process to correct fluctuation of expression;

an identifier generating step of generating an identifier uniquely specifying the document data or part thereof, based on all or part of the document data having been subjected to the canonicalization process in the canonicalization process step; and

an identity determining step of determining whether there is a common portion between one document data and another document data, based on the identifier having been generated in the identifier generating step.

3. An identity determining method comprising:

an identifier generating step of, based on all or

part of encoded data of document data, generating an identifier uniquely specifying the document data or part thereof; and

an identity determining step of determining
5 whether there exists a common portion between one document data and another document data, based on the identifier having been generated in the identifier generating step.

4. An identifier transmitting method comprising:
10 a canonicalization process step of subjecting document data to a canonicalization process to correct fluctuation of expression;

an identifier generating step of generating an identifier uniquely specifying the document data or
15 part thereof, based on all or part of the document data having been subjected to the canonicalization process in the canonicalization process step; and

an identifier transmitting step of transmitting the identifier having been generated in the identifier
20 generating step.

5. An identifier generating apparatus comprising:

canonicalization process means for subjecting document data to a canonicalization process to correct
25 fluctuation of expression; and

identifier generating means for generating an

identifier uniquely specifying the document data or part thereof, based on all or part of the document data having been subjected to the canonicalization process by the canonicalization process means.

5 6. The identifier generating apparatus according to Claim 5, wherein the canonicalization process means comprises type standardization process means for, using a class definition file of the document data describing a definition of a type, standardizing a type of
10 expression for a value described in the document data, in accordance with the type defined by the class definition file.

 7. The identifier generating apparatus according to Claim 6, wherein the type standardization process
15 means is configured to standardize an accuracy of numerical data described in the document data, in accordance with a definition of a type for numerical data described in the class definition file of the document data.

20 8. The identifier generating apparatus according to Claim 5, wherein the canonicalization process means comprises document data generating means for transforming first partial data and second partial data into document data in accordance with a predetermined
25 transformation rule.

 9. An identity determining apparatus comprising:

canonicalization process means for subjecting document data to a canonicalization process to correct fluctuation of expression;

5 identifier generating means for generating an identifier uniquely specifying the document data or part thereof, based on all or part of the document data having been subjected to the canonicalization process by the canonicalization process means; and

10 identity determining means for determining whether there is a common portion between one document data and another document data, based on the identifier having been generated by the identifier generating means.

15 10. The identity determining apparatus according to Claim 9, wherein the canonicalization process means comprises type standardization process means for, using a class definition file of the document data describing a definition of a type, standardizing a type of expression for a value described in the document data, in accordance with the type defined by the class
20 definition file.

25 11. The identity determining apparatus according to Claim 10, wherein the type standardization process means is configured to standardize an accuracy of numerical data described in the document data, in accordance with a definition of a type for numerical

data described in the class definition file of the document data.

12. The identity determining apparatus according to Claim 9, wherein the canonicalization process means
5 comprises document data generating means for transforming first partial data and second partial data into document data in accordance with a predetermined transformation rule.

13. The identity determining apparatus according to Claim 9, further comprising a cache preliminarily
10 storing the identifier having been generated by the identifier generating means, in correlation with the document data or a result of a predetermined process on the document data,

15 wherein the identity determining means is configured to:

perform a search inside the cache on the basis of the identifier of the document data as a target for a determination on identity, which has been generated by
20 the identifier generating means, determine that there exists identical document data, if the same identifier as the aforementioned identifier is present, and determine that there exists no identical document data, if the same identifier as the aforementioned identifier
25 is absent.

14. The identity determining apparatus according

to Claim 9, further comprising second identifier
generating means for generating an identifier uniquely
specifying the document data or part thereof, based on
all or part of the document data, prior to execution of
5 the canonicalization process by the canonicalization
process means,

wherein identity between one document data and
another document data is determined on the basis of the
identifier having been generated by the second
10 identifier generating means and wherein if they are
determined not to be identical, the canonicalization
process means performs the canonicalization process of
the document data.

15 15. An identity determining apparatus
comprising:

identifier generating means for, based on all or
part of encoded data of document data, generating an
identifier uniquely specifying the document data or
part thereof; and

20 identity determining means for determining
whether there exists a common portion between one
document data and another document data, based on the
identifier having been generated by the identifier
generating means.

25 16. The identity determining apparatus according
to Claim 15, wherein an instruction to skip a process

of decoding the encoded data of the document data is issued when the identity determining means determines that the two document data are identical.

5 17. An identifier transmitting apparatus comprising:

 canonicalization process means for subjecting document data to a canonicalization process to correct fluctuation of expression;

10 identifier generating means for generating an identifier uniquely specifying the document data or part thereof, based on all or part of the document data having been subjected to the canonicalization process by the canonicalization process means; and

15 identifier transmitting means for transmitting the identifier having been generated by the identifier generating means.

20 18. The identifier transmitting apparatus according to Claim 17, wherein the identifier transmitting means comprises transmitted data generating means for generating data obtained by replacing all or part of the document data by the identifier.

25 19. The identifier transmitting apparatus according to Claim 18, wherein the transmitted data generating means is configured to generate transmitted data described by an identifier uniquely specifying

partial data included in the document data, and difference data between the partial data and the document data.